

Estimation of Hydraulic Conductivity and Its Uncertainty from Grain-Size Data Using GLUE and Artificial Neural Networks

Bart Rogiers · Dirk Mallants · Okke Batelaan ·
Matej Gedeon · Marijke Huysmans ·
Alain Dassargues

Received: 3 August 2011 / Accepted: 30 April 2012
© International Association for Mathematical Geosciences 2012

Abstract Various approaches exist to relate saturated hydraulic conductivity (K_s) to grain-size data. Most methods use a single grain-size parameter and hence omit the information encompassed by the entire grain-size distribution. This study compares two data-driven modelling methods—multiple linear regression and artificial neural networks—that use the entire grain-size distribution data as input for K_s prediction. Besides the predictive capacity of the methods, the uncertainty associated with the model predictions is also evaluated, since such information is important for stochastic groundwater flow and contaminant transport modelling.

Artificial neural networks (ANNs) are combined with a generalised likelihood uncertainty estimation (GLUE) approach to predict K_s from grain-size data. The resulting GLUE-ANN hydraulic conductivity predictions and associated uncertainty

B. Rogiers (✉) · M. Gedeon
Institute for Environment, Health and Safety, Belgian Nuclear Research Centre (SCK•CEN),
Boeretang 200, 2400 Mol, Belgium
e-mail: bart.rogiers@sckcen.be

B. Rogiers · O. Batelaan · M. Huysmans · A. Dassargues
Dept. of Earth and Environmental Sciences, KU Leuven, Celestijnenlaan 200e - bus 2410, 3001
Heverlee, Belgium

O. Batelaan
Dept. of Hydrology and Hydraulic Engineering, Vrije Universiteit Brussel, Pleinlaan 2, 1050
Brussels, Belgium

A. Dassargues
Hydrogeology and Environmental Geology, Dept. of Architecture, Geology, Environment and Civil
Engineering (ArGEEnCo) and Aquapole, Université de Liège, B.52/3 Sart-Tilman, 4000 Liège,
Belgium

D. Mallants
Groundwater Hydrology Program, CSIRO Land and Water, Waite Road - Gate 4, Glen Osmond, SA
5064, Australia

estimates are compared with those obtained from the multiple linear regression models by a leave-one-out cross-validation. The GLUE-ANN ensemble prediction proved to be slightly better than multiple linear regression. The prediction uncertainty, however, was reduced by half an order of magnitude on average, and decreased at most by an order of magnitude. This demonstrates that the proposed method outperforms classical data-driven modelling techniques. Moreover, a comparison with methods from the literature demonstrates the importance of site-specific calibration. The data set used for this purpose originates mainly from unconsolidated sandy sediments of the Neogene aquifer, northern Belgium. The proposed predictive models are developed for 173 grain-size K_s -pairs. Finally, an application with the optimised models is presented for a borehole lacking K_s data.

Keywords Early stopping · Cross-validation · Generalised likelihood uncertainty estimation · Artificial neural networks · Sedimentary aquifer · Data-driven modelling · Likelihood measures · Principal component analysis · GLUE-ANN

1 Introduction

The quantification of saturated hydraulic conductivity (K_s ; see the list of symbols) is important for developing groundwater flow and contaminant transport models to support decision making about environmental impact assessment, groundwater contamination clean-up, water resources evaluation, or site monitoring. As an alternative for measuring K_s , grain-size data has often been used to indirectly obtain K_s , as it is one of the cheapest methods to obtain predictor variables for K_s . Hence, the development of (semi-)empirical methods to provide reliable K_s predictions has received lots of attention since the end of the 19th century. The large amount of literature devoted to developing pedotransfer functions in soil physics sciences indicates that measuring K_s directly is not always the most effective way to obtain the needed information, with additional arguments demonstrating the need for predicting K_s instead of measuring it. A first argument is that physically disturbed samples of subsurface sediments are much more common than undisturbed samples. Only good quality borehole cores allow for K_s measurements on minimally disturbed sediment. Grain-size analysis can however still be performed on disturbed core samples. A pedotransfer function developed in the previous setting gives the possibility to make detailed K_s estimates, with a corresponding uncertainty, when undisturbed samples are absent. Since disturbed sediment is easier and less expensive to retrieve than undisturbed borehole cores, this makes the characterisation of small-scale hydraulic conductivity much more efficient. A second argument is that grain-size data is definitely more ubiquitous than K_s data. If one thing is measured in the research sector or within the industry concerning unconsolidated sediments, it is the grain-size distribution. This is the most basic characteristic of the sediment, and is therefore very important to know when correlating different lithological units in space. The ability to make reliable predictions of K_s , starting from grain size, therefore allows for gathering much more information on the hydrogeology. This is especially true if former studies, or studies with different goals, were performed at a certain site or for a certain sediment.

List of symbols

Symbol	Explanation
p	Layer number
$i; j$	Neuron number
$w_{i,j}^p$	Weight for output of i th neuron in layer $p - 1$ to input for j th neuron in layer p
b_i^p	Bias i th neuron in layer p
x_i^p	Output i th neuron in layer p
y	Model predictions
Y	Observations
η	Learning rate
δ	Delta error
α	Momentum parameter
$\mu; d$	Learning rate adaptation factors
K_s	Saturated hydraulic conductivity
$C_h; C_{K-C}$	Hazen and Kozeny–Carman empirical coefficient
$L_2; L_{11}; \dots; L_{1000}$	Grain-size fractions $< 2; 11; \dots; 1000 \mu\text{m}$
f_i	i th grain-size fraction
$D_{li}; D_{si}$	Largest and smallest grain-size diameters of fraction f_i
e	Void ratio
SF	Grain shape factor
$s; S$	Sample number; total number of samples
$L(\theta_i Y)$	Likelihood for the i th parameter set conditioned on the observations Y
$\sigma_{\varepsilon,i}^2$	Error variance for the i th model
$\sigma_{\varepsilon'}^2$	Maximum likelihood estimate of the error variance
σ_{obs}^2	Observed variance
N	Weighting power
n	Number of samples used in model derivation
n_p	Amount of neurons in layer p

One of the earliest predictive K_s models is the Hazen Equation (Hazen 1892), which expresses the hydraulic conductivity (or permeability) as function of the square of the 10th percentile grain-size diameter (d_{10})

$$K_s = C_h d_{10}^2, \quad (1)$$

with the empirical coefficient C_h . This coefficient however ranges over three orders of magnitude in the literature (Carrier 2003). Numerous attempts have been made since Hazen to relate hydraulic conductivity and grain-size distribution for soil and sediments. In their review, Cronican and Gribb (2004) mention the use of the following grain-size parameters: d_{10} , geometric mean grain-size diameter (d_m), sand, silt and clay percentages, $d_{50}-d_{10}$, and the x -intercept of the $d_{50} \sim d_{10}$ straight line. Vienken and Dietrich (2011) recently evaluated several existing methods in which d_{20} , d_{50} , d_e (effective diameter) and the coefficient of uniformity $U = d_{60}/d_{10}$ are used. An attempt to correlate K_s to grain-size data for hydrogeological applications for clay-rich sediments is reported by Wemaere et al. (2008).

One of the main limitations of these (semi-)empirical formulations is that they all use one or very few parameters from the available grain-size distribution. While certain cumulative percentages or particle sizes might yield a reasonable correlation for a given hydraulic conductivity data set, the remaining information encompassed by the entire grain-size distribution is often omitted. A typical example of a predictive model that includes information from the entire grain-size distribution through the use of an effective particle diameter (d_e), is the Kozeny–Carman Equation (Kozeny 1927; Carman 1938, 1956)

$$K_s = C_{K-C} \left(\sum f_i / (D_{li}^{0.404} x D_{si}^{0.595}) \right)^{-2} (1/SF^2) (e^3 / (1 + e)), \quad (2)$$

with the empirical coefficient C_{K-C} , grain-size fractions f_i , with largest and smallest grain-size diameters D_{li} and D_{si} , a grain shape factor SF , and the void ratio e . Because there are still limitations associated with the latter model (Carrier 2003), approaches that allow calibration of the K_s versus grain-size distribution relationship have been developed. Van De Genachte et al. (1996) used principal component analysis in combination with multiple linear regression to model infiltration parameters, including K_s , of sandy soil. Another example of such development is the use of artificial neural networks (ANNs) for modelling soil hydraulic properties including K_s as function of grain-size distribution and possibly other material parameters such as bulk density, organic matter and carbonate content (Schaap et al. 1998; Pachepsky et al. 1999).

ANNs are computational tools that have found extensive application in a wide range of research areas and can be seen as a form of non-linear regression (Basheer and Hajmeer 2000). Moreover, it has been proven theoretically that multilayer feed-forward networks are universal approximators (Hornik et al. 1989; Hornik 1991). Environmental applications of this technique include stream flow, flood and rain-fall forecasting (Kişi 2007; Jain and Kumar 2007; Tiwari and Chatterjee 2010; Valverde Ramírez et al. 2005), processing remote sensing data (Linderman et al. 2004), flow and transport simulations (Morshed and Kaluarachchi 1998), ground-water table prediction (Coppola et al. 2005; Joorabchi et al. 2009) and water quality modelling (Zou et al. 2007; Khalil et al. 2011). In soil science, ANNs have been used extensively for predicting soil hydraulic properties from more easily measurable parameters. Schaap and Leij (1998) and Schaap et al. (2001) used ANNs for predicting soil water retention and log-transformed soil hydraulic conductivity from limited to extended sets of basic soil properties. The possible input variables encompass sand, silt and clay percentages, bulk density and retention points at 10 or 33 kPa. A hierarchical set of practical models available in the ROSETTA software was created for different levels of data availability (Schaap et al. 2001). Minasny et al. (2004) used sand, silt, clay content, bulk density, and saturated water content in an ANN prediction of soil hydraulic functions measured by multistep outflow data. Nakhaei (2005) used 8 cumulative particle size fractions for predicting log-transformed hydraulic conductivity for loamy sand soils with ANNs, with individual modelling of different soil types being found superior to joint modelling. Agyare et al. (2007) used in addition to grain size and bulk density, organic carbon and nine different terrain parameters to predict K_s .

As discussed above, most efforts of applying ANNs for predicting K_s are situated within the soil sciences, and use mainly clay, silt and sand content as grain-size input variables. There are fewer applications of the method that use data from sedimentary aquifers, possibly due to data scarcity which results from limited accessibility. Moreover, in the framework of stochastic modelling and risk assessment, the quantification of uncertainty related to these predictions is of equal importance, and this aspect seems to be addressed rarely. Uncertainty estimation is however not standard practice in ANN modelling. Asefa (2009) and Tiwari and Chatterjee (2010) show that using random input and network parameter selection in combination with traditional ANN modelling allows for obtaining uncertainty estimates. The framework that is developed here to produce in combination reliable predictions and their uncertainties, makes use of the generalised likelihood uncertainty estimation (GLUE) approach (Beven and Freer 2001). The principle of GLUE is based on a rejection of the concept of the optimal model. Instead, multiple models are considered to be equally acceptable if their performance, which is assessed through a likelihood measure, is alike. A quantitative measure of model performance is thus used to assess the acceptability of each model and to weigh each model accordingly in the model ensemble prediction and uncertainty estimates.

The objective of this paper is to develop an ANN ensemble modelling approach to predict hydraulic conductivity with multiple grain-size fractions as data input, and to provide reasonable uncertainty estimates using the GLUE methodology. The resulting GLUE-ANN hydraulic conductivity predictions and associated uncertainty estimates are compared with the ones obtained from linear regression models by using a leave-one-out cross-validation approach. An example application is shown in which hydraulic conductivity predictions are made for locations where only grain-size data was available. The K_s and grain-size data used originate from the Neogene aquifer in northern Belgium, which consists of several mostly unconsolidated sub-layers of quartz-rich sands with hydraulic conductivity values ranging between 10^{-3} and 10^{-11} m/s, the latter from clay lenses within an aquitard.

2 Methods and Materials

2.1 Measurements

2.1.1 Core Sampling

The study area covers approximately 60 km² in the north of Belgium and is part of the Nete basin (Fig. 1). At the study site the top 11–26 m of the aquifer are mainly composed of (i) Quaternary deposits (2–5 m thick) with grain-size varying from fine to coarse sand; (ii) well-sorted fine, white quartz sands usually referred to as Mol Sands that become more greenish medium sand in the western part of the study area; and (iii) underlying Kasterlee Sands that contain up to 1 weight % glauconite (Beerten et al. 2010). A second, much thicker aquifer exists under a very heterogeneous aquitard, the latter between 5 and 7 m thick and referred to as Kasterlee Clay. Stratigraphically the top of the second aquifer mainly consists of very coarse Diest Sands (approximately 84 m thick in the Dessel 5 borehole), with glauconite up to 30 weight %. The

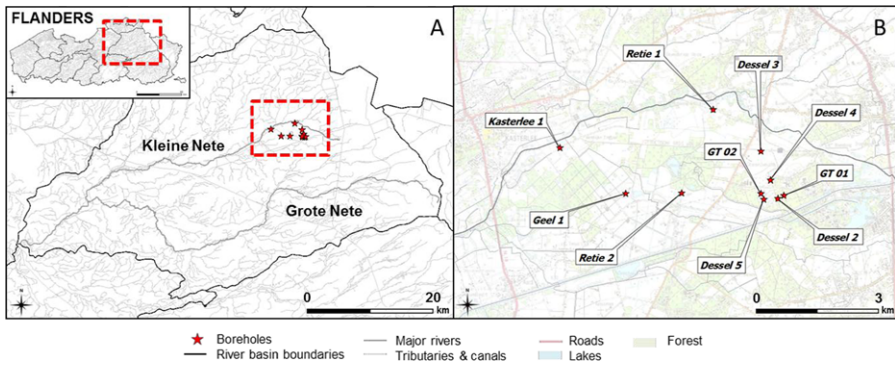


Fig. 1 (A) Location of the study area in the Nete basin and Flanders, and (B) the 10 borehole locations

Dessel, Berchem and Voort (DBV) sands are present below the Diest sands, overlying the gradual transition to the Boom Clay, which forms a regional aquitard (Wemaere et al. 2008). The total thickness of the DBV layers amounts to around 70 metres at the study site.

Nine boreholes were drilled to depths ranging from 50 (6 boreholes), 90 (2 boreholes) to 200 m. Contiguous coring was undertaken for all 50 and 90 m long cores and for the lower section of the 200 m long borehole (from 90 to 200 m). A pushing corer was used, developed for unconsolidated sediments, resulting in high recovery and low disturbances. The undisturbed cores were retrieved in 1.1 m long and 96 mm inner diameter PVC tubes and stored for further analyses (Beerten et al. 2010). The reconnaissance boreholes with a depth of 50 m (Dessel 3, Dessel 4, Geel 1, Kasterlee 1, Retie 1 and Retie 2) and 200 m (Dessel 5) were all used for model development, with hydraulic conductivity measurements being available for each of them. For the two 90 m deep boreholes (named GT 01 and GT 02), only grain-size data were available; they will be used in a final application of K_s prediction using the proposed GLUE-ANN model. Both reconnaissance and 90 m long boreholes penetrate the upper aquifer, the Kasterlee Clay aquitard, and the lower aquifer. Sampling consisted of sawing the 1.1-m long PVC cores longitudinally in two unequal parts (approximately in a 20/80 % division of the core diameter). The upper longitudinal section was used for geological description, while the lower part was used for sampling before being vacuum-packed for storage awaiting future analysis. Cylindrical plugs (100 cm³) were pressed in the cores for hydraulic conductivity (horizontal and vertical K_s) measurements and grain-size analysis (Beerten et al. 2010).

2.1.2 Grain-Size Measurements

Grain-size measurements were performed by means of two different techniques: a detailed analysis using a sedigraph yielding 20 size fractions (approximately 0.5, 0.7, 1, 1.4, 2, 3, 4, 6, 8, 11, 16, 22, 32, 44, 63, 88, 125, 177, 250 and greater than 250 μm), and a standard method (European standard EN 933-1) yielding 8 size fractions (2, 10, 20, 50, 100, 200, 500 and 1000 μm). The former technique was applied to clayey samples,

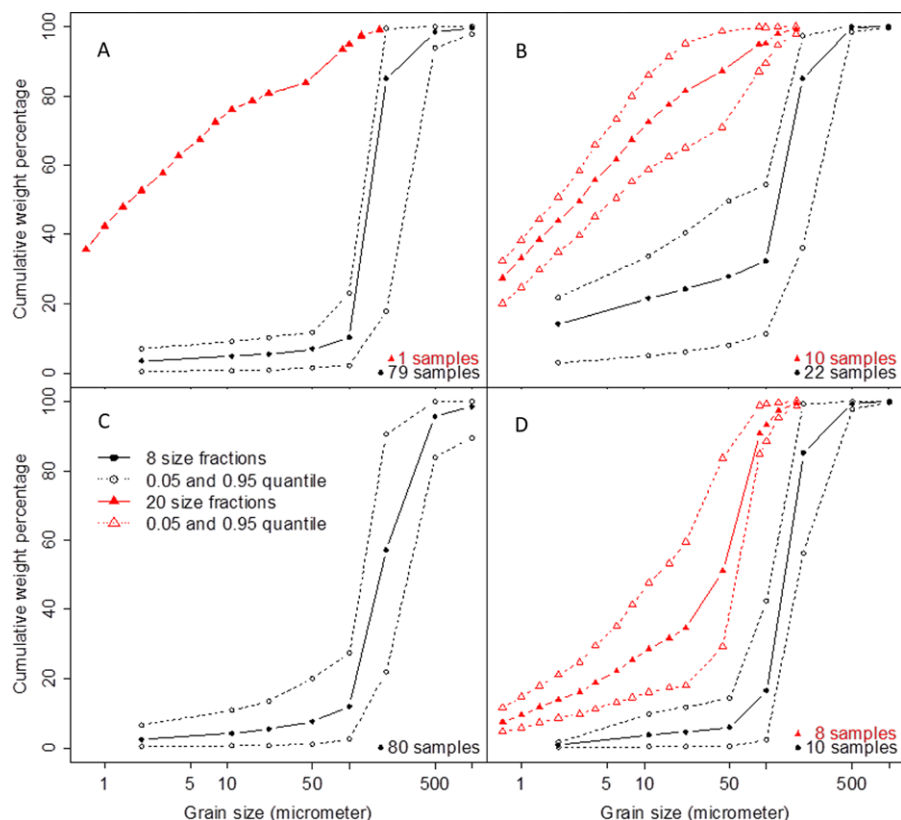


Fig. 2 Grain-size distributions of the upper aquifer (A), the Kasterlee Clay aquitard (B), the lower aquifer (C), and the deepest part (90–200 m) of the lower aquifer (D). For C only sand samples (8 size fractions) were analysed. In the other graphs, the clay samples (20 size fractions) were plotted separately

while the latter was applied to the sandy samples (see Fig. 2). Grain-size analyses with the sedigraph method were performed after eliminating carbonates by acidification, organic matter by oxygenation and particles larger than 250 μm by sieving. Characterisation of the particles is based on the sedimentation while the quantification is based on the attenuation of an X-ray beam. Because of the higher resolution in the fine particle ranges (from approximately 22 to 0.5 μm), it was specifically applied to the clayey samples (19 in total). Grain-size analyses of the sand samples have been performed according to the European standard EN 933-1. Samples were prepared by eliminating carbonates and organic matter as was done for the sedigraph method. For the greater than 50 μm particles, standard sieving was used with mesh sizes of 100, 250, 500, and 1000 μm . For the smaller than 50 μm particles, a peptising agent was added first after which the solid/water mix was put in a suspension cylinder for determination of the fine particles with class boundaries of 2, 11, and 22 μm . All particle fractions are expressed in mass percentage of the total dry weight, after removal of carbonates and organic matter. The reference diameters of the particle classes used in this study are thus: 2, 11, 22, 50, 100, 250, 500 and 1000 μm . The respective cumula-

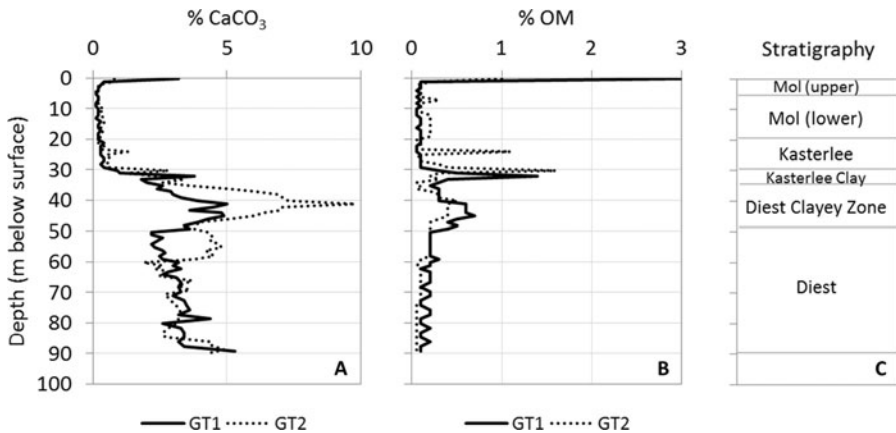


Fig. 3 (A) Carbonate content, (B) organic matter, and (C) stratigraphy for the two geotechnical boreholes (GT1 & GT2)

tive grain-size fractions are indicated as $L_2, L_{11}, \dots, L_{1000}$. Given that the analyses were performed using two different techniques whose interval diameters did not always coincide, intermediate fractions were obtained by linear interpolation of the lognormal distribution to allow both data sets to be combined in one analysis. The average cumulative grain-size distributions were calculated for each stratigraphical unit and are shown in Fig. 2.

The carbonate and organic matter content was checked for the entire stratigraphical succession because of the potential importance in determining K_s . Wang et al. (2009) demonstrated a negative correlation between hydraulic conductivity and organic matter, which is more important than grain size in the case of a very uniform, well-sorted soil. Hydraulic conductivity falls back to half or maximum one third of its initial value, with an increase in organic matter of 0.5 %. This paper deals however with considerably more heterogeneity, both within a single sample as across samples. Therefore, the role of organic matter is expected to be negligible. Organic matter in the soils occurring at the studied site is on average 6.4 % in the first metre below surface (Soetens 2008). However, the geological sediment subjects of this study are mainly part of the first 50 metres below surface. Organic material is hardly present (mostly below 0.3 %, never exceeding 2 %). The same is true for the carbonate content within the upper aquifer (below 0.5 %), but for the lower aquifer the average value amounts to 3 to 4 %, with an occasional peak between 5 and 10 %. Both the logs obtained from the two geotechnical boreholes are given in Fig. 3. Given these low percentages, it is assumed that the organic matter and carbonate content only have very little influence on the hydraulic conductivity in comparison with the textural information. The few samples with carbonate content exceeding 5 % might however be influenced significantly, and not accounting for this quantity might indeed induce errors. The lack of data for the entire data set did however not permit the inclusion of this parameter in the analysis, but the presented method can always be extended to include additional parameters.

2.1.3 Saturated Hydraulic Conductivity Measurements

A total of 280 hydraulic conductivity measurements were performed on cylindrical plugs of 100 cm³ using a constant-head permeameter based on the principles of Klute (1965). This was assumed to be more practical than a falling-head method, because of the amount of samples and the fixed-head method allows for a simpler laboratory set-up where several samples are measured at the same time. Given the hydraulic conductivities of the studied sediments, these fixed-head permeameter tests would have required a few tens of minutes up to a few days. Because of the very low gradients applied, the permeability of the samples is not believed to be altered. Additional 27 hydraulic conductivity measurements were performed using a modified constant-head permeameter adapted to medium-to-low K_s -values as described in Wemaere et al. (2002). For the latter, cylindrical samples are transferred into a stainless steel cylindrical cell with two sintered stainless steel filters (pore diameter of 10 μ m) at both ends. De-ionised water is injected at the bottom of the sample under a constant pressure of about 600 kPa and K_s is determined from flow measurements using Darcy's law (1856). The modified constant-head permeameter set-up is particularly suitable for clayey sediments with K_s in the range 10^{-13} to 10^{-9} m/s. For high clay contents, the measurement can last for three weeks until steady-state flow is achieved. In the case of more silty or sandy samples, the technique is slightly modified using a 1-m high water column with measurement times usually less than one day. The accuracy is about 8 %. Since we are dealing with Tertiary material, subject to groundwater flow for millions of years, the clay particles are assumed to be relatively immobile under such conditions. Microbial growth is not expected to occur within the time frame of the measurements.

About half of the samples were taken in a horizontal direction, whereas the other half were taken vertically, about 10 cm away from the horizontal sampling locations. The combined sample pairs were studied with statistical testing for the equality of the mean and variance of the obtained K_h and K_v data. On average, the ratio of geometric mean K_h and K_v values amounts to 2. However, the difference between the two data sets was tested using the statistical t -test and F -test. The p -values were respectively 0.104 and 0.076 for the upper aquifer, indicating that the difference of the mean and variance of the K_h and K_v populations is not significant at the 5 % level. For the aquitard and the lower aquifer, the t -test p -values amounted to 0.30 and 0.005, and 0.42 and 0.04 for the F -test. The difference between the K_h and K_v data sets thus is not statistically significant, except when the lower aquifer data set is considered separately. There, the presence of a few outliers might be the cause of this, as can be seen in the scatter plots of K_v versus K_h in Fig. 4. This is also indicated by the coefficient of determination values of 0.46, 0.67 and 0.38 for the upper aquifer, aquitard, and lower aquifer, respectively. In any case, it is very clear that the K_v values are not systematically lower than the corresponding K_h values, since 35 % of the paired samples show higher K_v values. Moreover, the average observed anisotropy is very small in comparison with the total K variability. Therefore, the data sets are considered to be compatible, and no difference is made between K_h and K_v samples.

Grain-size analyses were however only performed on the K_h samples. This means that the developed models will in principle be adequate for predicting only K_h data.

Fig. 4 Vertical versus horizontal K_s measurements in the three different hydrogeological units, with a linear model fit for each unit

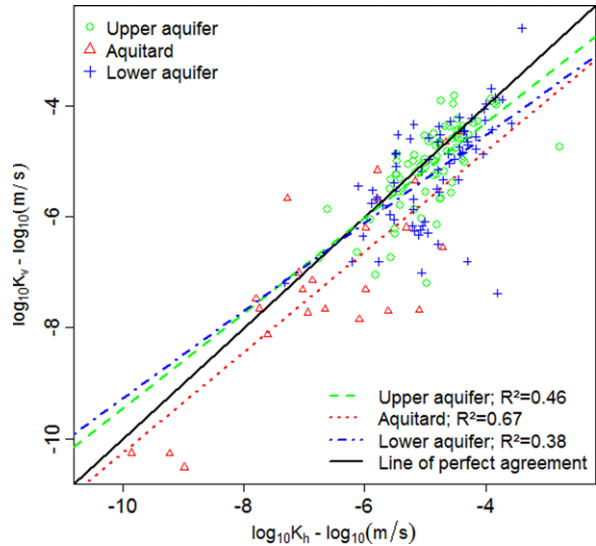
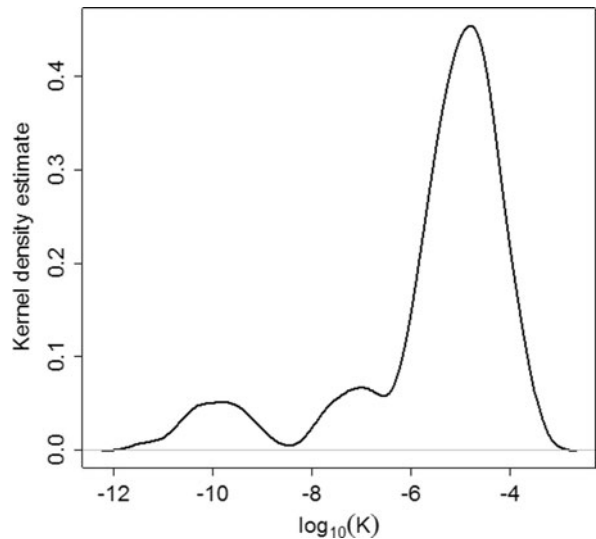


Fig. 5 Kernel density estimates of all 173 hydraulic conductivity samples used in the model derivation



However, because of the strong indications of isotropy at the sample scale, in this case the predictions will represent both the K_h and K_v values. A total number of 173 pairs of K_s -grain-size data were therefore available, and are used in the further analysis. The K_s kernel density estimate is shown in Fig. 5. This is a non-parametric way of estimating the probability density function of a random variable. One of the advantages of the smooth kernel density estimates compared to discrete histograms is their higher statistical efficiency (Scott 1979). Core samples for which only hydraulic conductivity was measured were used to determine the vertical anisotropy factor (Beerten et al. 2010).

Table 1 Correlations between grain-size fractions and \log_{10} -transformed hydraulic conductivity

Fraction	$\rho(\log_{10}K)$	$\rho(L_2)$	$\rho(L_{11})$	$\rho(L_{22})$	$\rho(L_{50})$	$\rho(L_{100})$	$\rho(L_{200})$	$\rho(L_{500})$	$\rho(L_{1000})$
L_2	-0.79	1.00	0.99	0.98	0.94	0.85	0.31	0.19	0.12
L_{11}	-0.84	0.99	1.00	1.00	0.97	0.89	0.31	0.18	0.11
L_{22}	-0.84	0.98	1.00	1.00	0.98	0.90	0.31	0.18	0.11
L_{50}	-0.9	0.94	0.97	0.98	1.00	0.96	0.32	0.17	0.11
L_{100}	-0.93	0.85	0.89	0.90	0.96	1.00	0.34	0.17	0.10
L_{200}	-0.43	0.31	0.31	0.31	0.32	0.34	1.00	0.57	0.41
L_{500}	-0.22	0.19	0.18	0.18	0.17	0.17	0.57	1.00	0.84
L_{1000}	-0.14	0.12	0.11	0.11	0.11	0.10	0.41	0.84	1.00

2.2 Modelling

2.2.1 Linear Model

All cumulative grain-size fractions ($L_2, L_{11}, \dots, L_{1000}$) were used as predictor variables. These are however highly correlated, as is indicated in Table 1. Regressing on multi-collinear predictor variables has a detrimental effect on regression analysis and has to be avoided (Gunst and Mason 1980). Therefore, multiple regression was applied after orthogonalisation of the data using principal components analysis (PCA) as recommended by Van De Genachte et al. (1996). PCA produces orthogonal, uncorrelated principal components as new independent variables that are linear combinations of the original variables (Hill and Lewicki 2007). PCA was performed on the covariance matrix, therefore putting more weight on the most variable fractions, because the performance of correlation-based PCA was slightly lesser. The component loadings and correlation coefficients are given in Table 2. By loadings it is meant the weight by which each original variable should be multiplied to get the component score. The columns of Table 2 equal the eigenvectors of the covariance matrix. The first PC explains 76 % of the variance, the first two PCs explain 96 % of the variance, and the first three PCs explain 99 % of the variance. The largest diameters seem to carry very little information on the variability within the grain-size data, since they only load on a few components. Moreover, they show a very small correlation with K_s , as is shown in Table 1.

Subsequently, stepwise multiple linear regression (MLR) was performed with the principal component scores. The components with the highest explained variance were added first, and their significance was checked before adding another. All eight PCs were evaluated, even though the higher order PCs accounted for a very small percentage of the variance. This yielded a regression model based on the first (adjusted $R^2 = 0.836$), third (0.878), fifth (0.882) and eighth PC (0.888). The contribution of the others PCs in the linear regression is statistically insignificant. The first PC shows negative weights for all grain-size fractions, all being more or less equally important. The third PC has a large positive contribution of L_{100} , while PC5 gets a negative contribution from L_2 and positively from L_{50} . Finally, adding PC8, with contributions from L_2, L_{11} , and L_{22} , to the regression model was still statistically significant.

Table 2 Variable loadings on the principal components of the grain-size data set, and the correlation coefficients (in parentheses)

Size fraction	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8
L_2	-0.22 (-0.94)	-0.06 (-0.13)	-0.32 (-0.28)	0.00 (0.00)	-0.56 (-0.13)	0.68 (0.08)	0.02 (0.00)	-0.24 (-0.01)
L_{11}	-0.38 (-0.97)	-0.11 (-0.14)	-0.41 (-0.21)	0.01 (0.00)	-0.24 (-0.03)	-0.26 (-0.02)	0.00 (0.00)	0.75 (0.01)
L_{22}	-0.43 (-0.97)	-0.13 (-0.15)	-0.41 (-0.19)	0.02 (0.01)	0.06 (0.01)	-0.51 (-0.03)	0.00 (0.00)	-0.61 (-0.01)
L_{50}	-0.49 (-0.99)	-0.14 (-0.14)	-0.03 (-0.01)	0.02 (0.00)	0.73 (0.08)	0.44 (0.03)	-0.01 (0.00)	0.10 (0.00)
L_{100}	-0.56 (-0.96)	-0.12 (-0.10)	0.75 (-0.26)	-0.04 (-0.01)	-0.31 (-0.03)	-0.10 (-0.01)	0.00 (0.00)	-0.03 (0.00)
L_{200}	-0.25 (-0.45)	0.96 (0.89)	-0.02 (-0.01)	0.12 (0.02)	0.01 (0.00)	0.00 (0.00)	0.01 (0.00)	0.00 (0.00)
L_{500}	-0.03 (-0.25)	0.11 (0.53)	-0.05 (-0.09)	-0.90 (-0.80)	0.02 (0.01)	0.00 (0.00)	-0.41 (-0.10)	0.00 (0.00)
L_{1000}	-0.01 (-0.16)	0.04 (0.39)	-0.02 (-0.07)	-0.41 (-0.78)	0.03 (0.04)	-0.01 (-0.01)	0.91 (0.45)	0.00 (0.00)
Cumulative variance accounted for (%)	76	96	99	99.68	99.89	99.935	99.997	100

Table 3 Performance measures (MSE = mean square error, MAE = mean absolute error, ME = mean error, R^2 = coefficient of determination) for the MLR and GLUE-ANN models. Two indicators for the uncertainty envelope around the ensemble estimate are: M95 % PR (= mean 95 % prediction interval range in $\log_{10} K$ space) and 95 % PO (= 95 % prediction interval outlier fraction)

Method	MSE	MAE	ME	R^2	M95 % PR	95 % PO
Modelling all data						
MLR	0.3	0.42	0.00	0.89	2.23	0.06
GLUE-ANN	0.21	0.34	-0.04	0.93	1.86	0.07
Leave-one-out cross-validation						
MLR	0.34	0.44	0.00	0.88	2.24	0.07
GLUE-ANN	0.29	0.40	-0.03	0.89	1.80	0.08

A back-transformation is needed to express the regression models in terms of the original particle size parameters. The resulting linear model is presented in Eq. (3), and some goodness-of-fit measures are presented in Table 3

$$\begin{aligned}
 \log_{10} K_s = & -3.42 + 1.4x10^{-1}L_2 - 3.26x10^{-1}L_{11} + 2.76x10^{-1}L_{22} \\
 & - 9.43x10^{-2}L_{50} - 1.96x10^{-2}L_{100} - 9.23x10^{-3}L_{200} \\
 & - 1.44x10^{-3}L_{500} - 2.04x10^{-3}L_{1000}.
 \end{aligned} \quad (3)$$

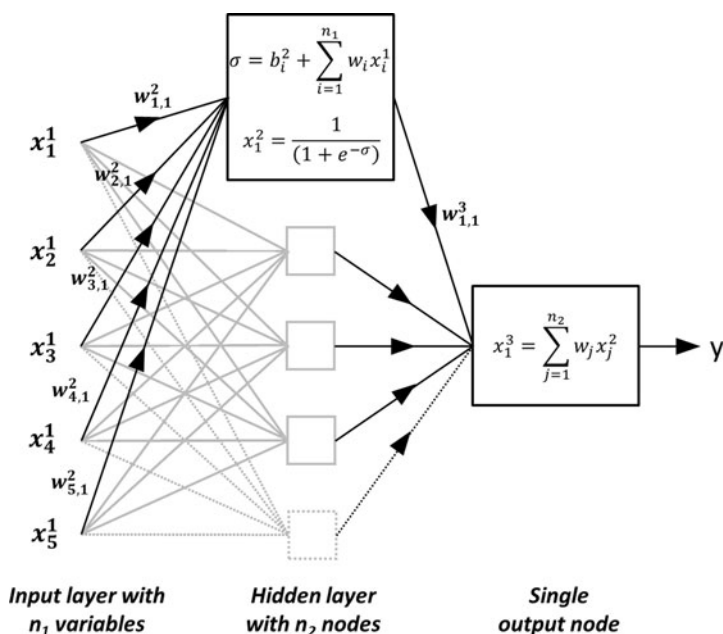


Fig. 6 Neural network architecture, weights, and activation function

The 95 % prediction intervals of this linear model are used to assess the uncertainty about the predicted K_s values (Kleinbaum et al. 2007).

2.2.2 GLUE-ANN Approach

Artificial Neural Networks Artificial neural networks consist, like their biological analogues, of several processing elements, called nodes or neurons, that are interconnected (Fig. 6). In a feedforward ANN architecture, information always moves in one direction; it is the simplest type of ANN as there are no loops in the network. It has input nodes, where the input data are entered, one or multiple so-called hidden layers containing a certain amount of nodes, and one or multiple output nodes, where the target variable predictions are provided. The number of hidden nodes is a critical parameter of any feedforward ANN: too many nodes may cause overfitting the data, thus resulting in poor generalisation on data not used for training. On the other hand, too few hidden nodes will cause underfitting of the model, which will therefore be insufficiently accurate. Most of the mathematics behind the connections and processing within the nodes are in fact linear combinations. The n_1 input variables (x_i^1) are recombined with weights ($w_{i,j}^2$) and a bias (b_i^2) to a set of n_2 new variables (or nodes) in the so-called hidden layer. The outputs of the hidden layers are recombined in a similar way, to provide the networks output. The amount of nodes in the hidden layer is not a fixed value, and depends on the complexity of the problem (Basheer and Hajmeer 2000). The optimum number of nodes also increases proportionally to the number of observations (Fletcher et al. 1998). The non-linear approximation capabilities of an ANN are a consequence of introducing so-called activation functions

in each of the networks nodes. Sigmoid functions are the most commonly used activation functions for the hidden layer, and are applied in this study. For the output node, however, a linear activation function was used.

Since the correlations between the predictors and the dependent variable are considerably high (Table 1), it is chosen to explain the residuals of the MLR model through non-linear modelling with the ANN approach. Different transformations of the predictor variables were tested for further analysis, but only the one with the best results is retained in this paper: a square root transform. In addition, standardising the inputs can make neural network training faster and reduces the chances of getting stuck in local optima (Sarle 1997). Therefore, the cumulative grain-size fractions were standardised to zero mean and unit standard deviation. Furthermore, to avoid difficulties with the training algorithm in performing efficient optimisation of the network biases and weights with highly correlated variables (see Table 2), orthogonalisation by PCA was performed similarly to the multiple regression analysis. Because of the universal approximation theorem (Hornik et al. 1989; Hornik 1991) and the continuous character of input and output variables, it was chosen to use the standard feedforward neural network architecture with one hidden layer to develop the model.

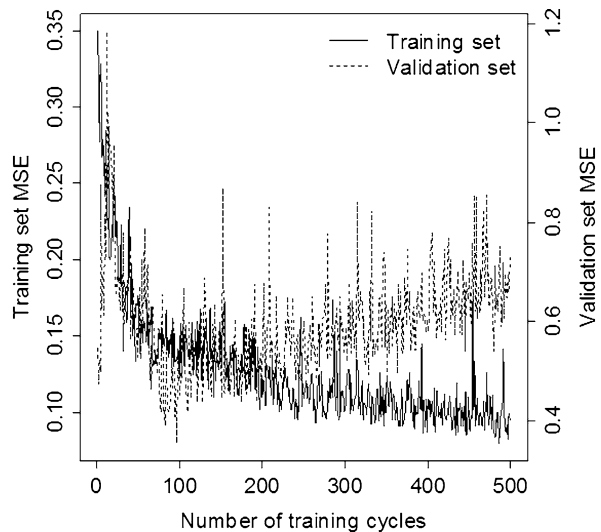
The adaptive gradient descent with momentum training method (Tollenaere 1990; Sarkar 1995; Haykin 1999), also known as SuperSAB, was used to calibrate the network weights and biases. This method is a modification of the standard gradient descent back-propagation algorithm that allows for a faster learning process. It is a combination of the momentum and the adaptive back-propagation methods. The momentum method includes a fraction of the last step in the parameter space when adjusting the network weights. Equation (4) shows how the neuron weight $w_{i,j}^p(t)$ for training cycle t is calculated based on the partial derivative of the objective function to the net target neuron input δ_i^{p+1} (delta error δ ; for a full derivation of the generalised delta rule, the reader is referred to Henseler 1995), times the output of the neuron that has to be weighted y_j^p , times the learning rate η , plus the weight of the former step $w_{i,j}^p(t-1)$ times the momentum parameter α . This causes the optimisation to move faster towards the minimum. However, high gradients around the minimum of the objective function might cause the algorithm to jump across. To solve this problem, the momentum method is combined with an adaptive back-propagation, which allows for adjusting the learning rate η according to changes of the sign of the objective function gradients. The learning rate η is increased if the sign of the gradient remains the same, or decreased if it changes during training by the factors μ or d , respectively, and is calculated by Eq. (5). The initial learning rate η and the momentum parameter α were selected by trial and error, and are equal respectively to 0.01 and 0.5. The same method was applied for the network biases, and the least mean square error (MSE) criterion was used to evaluate the model performance during training

$$\Delta w_{ij}^p(t) = -\eta \delta_i^{p+1} y_j^p + \alpha \Delta w_{ij}^p(t-1), \quad (4)$$

$$\eta_i^p(t) = \begin{cases} \mu \eta_i^p(t-1) & \text{if } \Delta w_i^p(t) \Delta w_i^p(t-1) \geq 0, \\ \delta \eta_i^p(t-1) & \text{if } \Delta w_i^p(t) \Delta w_i^p(t-1) < 0. \end{cases} \quad (5)$$

To ensure generalisation capacity of the trained neural network, the early stopping technique was invoked (Prechelt 1998). The network performance is checked

Fig. 7 Behaviour of ANN model performance (mean square error) as function of training cycles. When the MSE no longer decreases, an optimal number of training cycles has been reached and early stopping is recommended: at about 100 training cycles, the maximum generalisation capacity is reached. For larger training cycles, the network starts to become overtrained and the validation error rises again



regularly on a validation set during training. When the model performance starts to decrease, the network is being overtrained, and generalisation cannot be guaranteed. Hence, the weight configuration with the best performance is selected. Figure 7 shows an example of the evolution of the training and early stopping MSE as function of the number of iterations. With the early stopping technique, the network weights at about 100 training cycles are chosen in this example because of their maximum generalisation capacity (minimum validation MSE). For a higher number of training cycles, the network becomes overtrained, indicated by the increasing validation set MSE. One third of all samples were selected randomly for early stopping cross-validation. The implementation of the discussed ANN training algorithms by Castejón Limas et al. (2010) in the AMORE package for the R language (R Development Core Team 2010) was used in this paper.

The parameter that remains to be determined is n_2 : the amount of hidden nodes. Since the optimal number is dependent on the complexity of the problem amongst other things (Basheer and Hajmeer 2000), it is generally determined by trial and error. Several rules of thumb do exist in the literature, for example the number should not exceed twice the number of inputs (Swingler 1996; Berry and Linoff 1997), and approaches are available to perform a stepwise optimisation (Fletcher et al. 1998). We choose to select n_2 randomly with a maximum value of three times n_1 , the number of input variables. The early stopping procedure will still prevent a network with a large number of hidden nodes to become overtrained, and hence the uncertainty associated with n_2 will be accounted for.

Uncertainty Estimation by GLUE To obtain uncertainty estimations for the predicted K_s values, a Monte Carlo data set was generated by considering the following random variations in model parameters and data sets: (i) the number of hidden nodes (n_2); (ii) the initialisation of network weights ($w_{i,j}^p$) and biases (b_i^p); (iii) dimensionality reduction of the input data to a random number of input variables (n_1)

after principal component analysis; and (iv) the selection of early stopping data. A total of 10,000 Monte Carlo realisations were used to achieve convergence of the estimated uncertainty. This gives rise to an ensemble of 10,000 ANN models, which are all a priori equally likely. To update the likelihoods of the individual models, the concept of GLUE (Beven and Freer 2001) was used.

GLUE is based on a rejection of the concept of the optimal model. Instead, multiple models are considered to be equally acceptable if their performance is alike. This concept is also called equifinality. For the application of the GLUE method to this model ensemble, a number of choices have to be made explicitly (Beven and Freer 2001). First, a quantitative measure of model performance has to be used to assess the acceptability of each model and to weigh each model accordingly in the model ensemble prediction and uncertainty estimates. This measure is called the likelihood measure. A second choice involves the rejection of the lesser performing part of the model ensemble by selecting a behavioural threshold. In a final step, the model ensemble predictions are calculated by taking the weighted mean, median and 2.5–97.5 % weighted quantiles from the ensemble of behavioural model predictions, with the individual model likelihoods $L(\theta_i|Y)$ used as weights, possibly rescaled by the behavioural threshold. Considerable controversy exists however about the use of such subjective likelihood measures, but examples exist that are consistent with classical and Bayesian statistics. The approach proposed by Stedinger et al. (2008) was adopted here, which is valid for the assumption of normal and independently distributed model errors. Equation (6) is used as the likelihood measure, which equals the model weights

$$L(\theta_i|Y) = \kappa \exp[-n\sigma_{\varepsilon,i}^2/2\sigma_{\varepsilon'}'^2], \quad (6)$$

where $L(\theta_i|Y)$ is the likelihood for the i th parameter set, θ_i , conditioned on the observations Y , $\sigma_{\varepsilon,i}^2$ is the mean square error for the model with the set of parameters θ_i , $\sigma_{\varepsilon'}'^2$ is the maximum likelihood estimate of the error variance, and κ is a normalising constant. This ensures that (i) the best fitting model determines the standard at which other solutions are compared; (ii) the probability of each parameter set reflects how well the model matches the calibration data; and (iii) the number of samples n used in model derivation is accounted for in determining model performance.

When it comes to applications with the GLUE-ANN approach, for predicting K_s and its uncertainty for example, the following extension based on Stedinger et al. (2008) was implemented. These authors argue that the uncertainty interval of conventional GLUE analysis does not correspond to statistically sound prediction intervals, since the model error that describes the likely difference between the observations and their mean values is generally ignored. Using the likelihood measure in Eq. (6), one can easily post-process Monte Carlo results to account for model error, and end up with proper prediction intervals. This is accomplished by adding 50 different and independent zero-mean random normal variates with variance $\sigma_{\varepsilon'}'^2$ to each of the model outcomes, as suggested by Stedinger et al. (2008). The number of results in the model ensemble used for constructing 95 % prediction intervals thus increased to $50 \times 10,000$. The ensemble of 10,000 ANN models was used that way for generating $50 \times 10,000$ predictions for one of the 90 m long boreholes, for which grain-size data was available but not K_s .

2.2.3 Model Performance Assessment

To assess the performance of the developed models, a leave-one-out cross-validation was performed. Due to the limited number of samples, a simplified form of cross-validation where the observations are split into two sets—one for learning and one for control—has not been pursued. For each K_s out of the 173 samples, the model ensemble consisted of 1000 ANN models, trained with the other 172 samples. The number of models was set to 1000 to balance reasonable convergence of the estimated distributions and computational load (two orders of magnitude higher than the normal modelling). For enabling proper comparison with the linear regression, the same leave-one-out cross-validation was applied to the MLR model. The advantage of this approach is that no data should be kept apart to serve as a final test set after model derivation, and hence it is objective. It does require some additional computational time, but it is critical to the success of data-driven modelling methods that rely heavily on the amount of available data, as is the case for ANNs. Based on the R language, wrapper functions around the AMORE package algorithms (Castejón Limas et al. 2010) for a GLUE-ANN analysis are provided in the web page as linked to in the [Appendix](#). Several options are provided, including the leave-one-out cross-validation test using different likelihood measures, and a preprocessing routine for the input data. A small script is added to illustrate the use of these functions with a simple example.

3 Results

3.1 Predictive Power for K_s Models

First, the performance of the MLR model (Eq. (3)) and the optimised GLUE-ANN model are compared. The prediction versus observation scatter plots of the model ensemble based on all available samples and the model testing by leave-one-out cross-validation are given in Fig. 8 and Fig. 9, respectively. Overall, both models perform reasonably well, with all predictions within two orders of magnitude of the observations, and 90 % within one order of magnitude. The predictions of the GLUE-ANN model in the K_s range of 10^{-4} to 10^{-6} m/s appear to be less biased. Based on the performance measures provided in Table 3, the GLUE-ANN is demonstrated to perform better than the MLR according to all measures, except for the mean error. The difference in performance between both models is small, however.

Figure 10 shows the prior and posterior distributions of the number of hidden nodes in the model ensemble. As described before, the number was chosen randomly between one and three times the number of variables used as input for the network, and thus has a uniform prior distribution. Comparing this with the posterior distribution, given by applying the weights based on Stedinger et al. (2008), it becomes apparent that the optimal number of hidden nodes is 5 (which is also the architecture for the model with maximum likelihood). Networks with more than 12 hidden nodes seem to perform worse, and are thus penalised in the behavioural weighting. Some contribution comes from the models with about 9 hidden nodes.

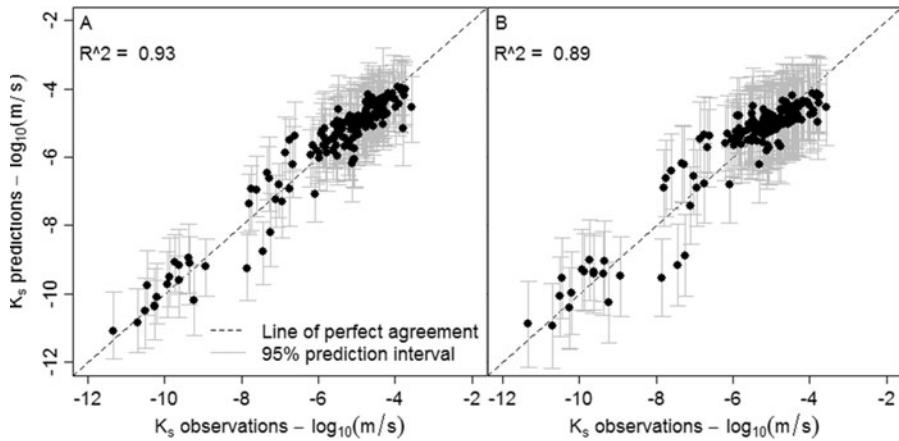


Fig. 8 Hydraulic conductivity observations versus model predictions and uncertainty estimates (95 % prediction interval) for the GLUE-ANN model (A), and the MLR model (B)

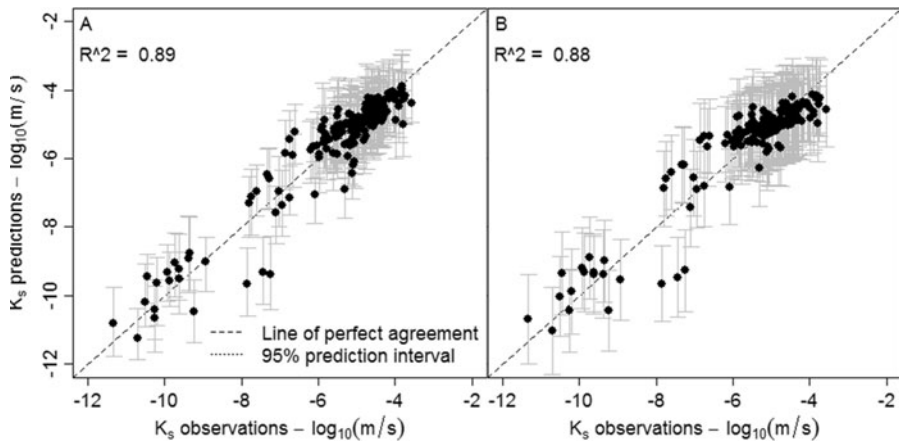


Fig. 9 Hydraulic conductivity observations versus model predictions and uncertainty estimates (95 % prediction interval) from the leave-one-out cross-validation for the GLUE-ANN model (A), and the MLR model (B)

To demonstrate the superior behaviour of predictive K_s models based on using multiple grain-size variables or the entire grain-size distribution compared to single parameter models, linear regression was also performed with the highest correlated grain-size fraction L_{100} . The MSE obtained with this model was 0.38, and increased to 0.48 with the leave-one-out cross-validation. Comparing these values with the model performances in case of MLR (Table 4) indicates that using a single grain-size fraction is 19 % less accurate than using the entire grain-size distribution (based on the ratio of the root MSE). The difference in accuracy with the non-linear ANN model increases to about 29 %.

Fig. 10 Kernel density estimate of the prior and posterior distributions of the number of hidden nodes

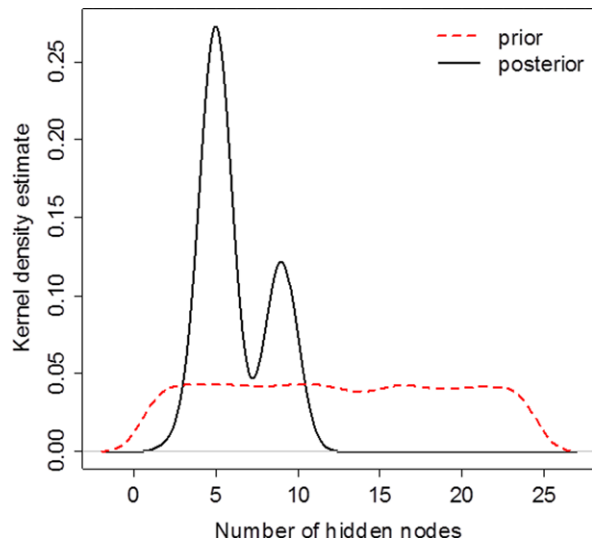


Table 4 Performance measures (MSE = mean square error, MAE = mean absolute error, ME = mean error, R^2 = coefficient of determination) for some literature K_s estimation methods discussed by Vienken and Dietrich (2011), and the predictive model of Rosetta (Schaap et al. 2001)

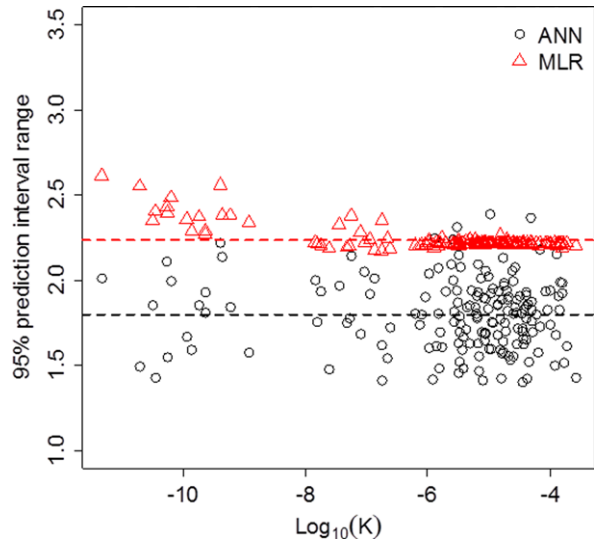
Performance measure	Hazen	Beyer	Kozeny–Köhler	USBR	Seelheim	Kaubisch	Terzaghi	Rosetta
MSE	1.49	9.43	0.89	0.8	2.57	1.71	1.71	2.44
MAE	0.93	2.9	0.74	0.63	1.23	1.05	0.97	1.09
ME	0.52	2.86	−0.42	0.23	1.21	0.99	0.26	0.99
R^2	0.63	0.62	0.73	0.72	0.62	0.71	0.64	0.54

Next, Table 4 provides an overview of the performance of several literature-based models to estimate K_s that are discussed by Vienken and Dietrich (2011), and the predictive K_s model of the Rosetta software by Schaap et al. (2001). A comparison between these values and the ones for the MLR and GLUE-ANN model indicate the benefit of using a site-specific calibration. Most of them perform worse on the MSE criterion than on the R^2 value, indicating that relative hydraulic conductivities can be obtained using these literature methods, but correct absolute values are more difficult to obtain without calibration. For example, based on the MSE, the USBR performs four times worse than the ANN model, while the Beyer model is more than 40 times worse.

3.2 GLUE-Based Prediction Uncertainty

To assess the uncertainty about the estimated hydraulic conductivity obtained by both MLR and ANN methods, the 95 % prediction intervals are compared between models. The average 95 % prediction interval in $\log_{10}(K_s)$ -space is included in Table 3, as well as the percentage of observations that fall outside the 95 % prediction intervals.

Fig. 11 Comparison of 95 % prediction intervals for the leave-one-out cross-validation



The GLUE-ANN approach appears to reduce the uncertainty about the K_s -values by half an order of magnitude, on average. This is also illustrated in Fig. 11, by plotting the 95 % prediction interval for all samples versus the observed hydraulic conductivity. The mean values are indicated by the horizontal dotted lines. The mean uncertainty owing to the GLUE-ANN model is about half of one order of magnitude smaller than that for the MLR model. Also note that for GLUE-ANN the uncertainty is consistent for the entire range of K_s -values; for MLR the uncertainty increases significantly for $\log_{10}(K_s) > -8$.

To illustrate the very different character of the hydraulic conductivity probability density functions obtained by both models, the distributions for a selected sample are shown in Fig. 12. The difference between the non-linear and linear methods is very clear. Multiple separate high-density regions can be present in the GLUE-ANN prediction, while the MLR model provides a Gaussian density function. In this particular case, even the third highest peak, which corresponds to the observed value, has a higher density than the MLR estimate, and is thus more likely. Evidently, the MLR model also outperforms the GLUE-ANN model in several other cross-validation cases, but overall, the estimation and uncertainty of the GLUE-ANN model are superior.

3.3 Application: Predictions of K_s and Its Uncertainty for a 90 m Deep Borehole

As illustration of the derived site-specific data-driven multi-parameter models, K_s predictions are made for an additional borehole, for which only grain-size data was available. Figure 13 shows the K_s predictions for the entire stratigraphical section. It seems that the GLUE-ANN K_s predictions are considerably higher in the Kasterlee unit compared to the Mol Lower unit. This is consistent with data from the Dessel 4 borehole (not shown), although the spatial resolution of the latter is only half that of the GT-01 borehole. Since the 90-m long borehole offers at least twice the resolution

Fig. 12 Comparison between the predictions of both models for a selected sample from the leave-one-out cross-validation, illustrating the benefit of the non-linear ensemble modelling approach

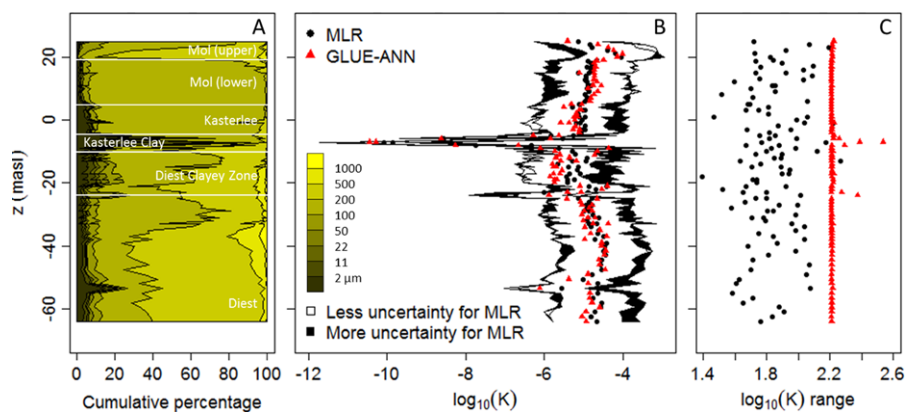
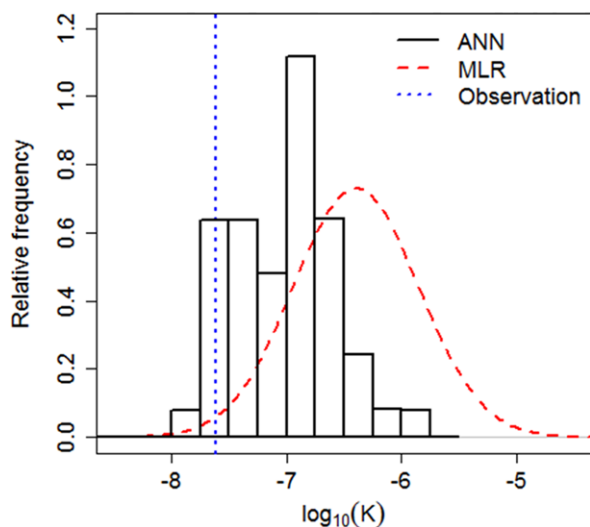


Fig. 13 (A) Grain-size data from a 90-m deep borehole (GT 01). (B) Application of the GLUE-ANN and multiple linear regression models for K_s predictions using data. The 95 % prediction intervals are plotted at both sides of the K_s estimates. If the area in between is coloured black, the side closest to the estimate (with the least uncertainty) represents the the GLUE-ANN model. The outer side represents the MLR model. If the area in between has a white colour, the MLR model has the least uncertainty. (C) Magnitude of the 95 % prediction interval for both the MLR and GLUE-ANN models

of the Dessel 4 reconnaissance borehole in terms of grain-size data, the greater detail in predicted K_s supports the use of systematically lower K_s -values in the Kasterlee unit for future groundwater modelling. For the lower aquifer (Diest Clayey Zone and Diest), the predictions of both models seem to be more similar besides some systematic lower values in the most upper and lowest parts. Several layers might now be distinguished in this lower aquifer, since the other boreholes were less deep. For instance, a slightly more conductive layer of about 10 m thickness can be found at the top of the Diest unit.

4 Conclusions

It was shown that using grain-size distributions together with linear and non-linear data-driven modelling gives strong improvement over traditional non-site-specific methods for predicting K_s . Multiple linear regression (MLR) combined with principal component analysis is found to be reasonably accurate for predicting hydraulic conductivity on the basis of grain-size data. Up to eight cumulative grain-size fractions were included in the analysis; only two fractions (L_{500} and L_{1000}) hardly contributed to explaining variability in the K_s data. This shows that nearly the entire particle-size distribution curve contains useful information to be included to build predictive K_s models. This was further demonstrated on the basis of a comparison with a linear regression model using as a single predictor variable the grain-size fraction with the highest correlation (L_{100}) with K_s . Considering this single grain-size fraction increases the error by 19 and 29 % respective to the MLR and GLUE-ANN model performances.

The non-linear data-driven modelling was performed by combining artificial neural networks (ANN) and generalised likelihood uncertainty estimation (GLUE). This GLUE-ANN approach provided a greater accuracy in predicted K_s , and considerably smaller prediction intervals, with equal reliability. The GLUE-ANN model is thus preferred over the MLR model, especially in the framework of site-specific stochastic groundwater flow modelling where small-scale variability in K_s is important. To validate the developed models, a leave-one-out cross-validation was performed. The advantage of this approach is that the maximum amount of data remains available for training the ANNs. The GLUE-ANN model proved to be more accurate than the MLR model, and the 95 % prediction uncertainty range was reduced on average by half an order of magnitude.

The extension of the proposed model, using additional parameters as organic matter and carbonate content, soil porosity and density could potentially further increase the performance of the model. A hierarchical set of models would then be useful to deal with different data availability (generic, limited site-specific, and detailed site-specific) at different locations. The results obtained from the textural information only are however already satisfying, and show a reasonable uncertainty. Other missing parameters can be added to the network output layer as well, in order to make predictions for all parameters, at locations with different parameter availability. The application of the methods to a 90-m long borehole revealed that the increased accuracy of the GLUE-ANN can highlight better the subtle differences between several geological units in terms of K_s . Moreover, the uncertainty associated with the GLUE-ANN estimates is for several stratigraphic units considerably less than the uncertainty on the MLR estimate. The applicability of this site-specific model to other sites remains however limited, although this was not tested. A large range of K_s -values was however used in the model derivation, therefore the application to sites with a similar geological setting may still be successful.

Acknowledgements The authors are grateful to ONDRAF/NIRAS, the Belgian Agency for Radioactive Waste and Enriched Fissile Materials, for providing the data. Findings and conclusions in this paper are those of the authors and do not necessarily represent the official position of ONDRAF/NIRAS.

Appendix—Web Page for a GLUE-ANN Analysis

<http://www.sckcen.be/en/content/download/16945/225528/file/GLUE-ANN-analysis.html>

References

- Agyare WA, Park SJ, Vlek PL (2007) Artificial neural network estimation of saturated hydraulic conductivity. *Vadose Zone J* 6(2):423–431
- Asefa T (2009) Ensemble streamflow forecast: A GLUE-based neural network approach. *J Am Water Resour Assoc* 45(5):1155–1163
- Basheer IA, Hajmeer M (2000) Artificial neural networks: Fundamentals computing design and application. *J Microbiol Methods* 43(1):3–31
- Beerten K, Wemaere I, Gedeon M, Labat S, Rogiers B, Mallants D, Salah S, Leterme B (2010) Geological hydrogeological and hydrological data for the Dessel disposal site Project near surface disposal of category A waste at Dessel—Version 1, NIROND-TR 2009-05 E, 261
- Berry M, Linoff G (1997) *Data mining techniques for marketing sales and customer support*. Wiley, New York, 464 pp
- Beven K, Freer J (2001) Equifinality data assimilation and uncertainty estimation in mechanistic modelling of complex environmental systems using the GLUE methodology. *J Hydrol* 249(1–4):11–29
- Carman PC (1938) The determination of the specific surface of powders. *J Soc Chem Ind Trans* 57:225–234
- Carman PC (1956) *Flow of gases through porous media*. Butterworths Scientific Publications, London
- Carrier WD (2003) Goodbye Hazen; hello Kozeny–Carman. *Journal of Geotechnical and Geoenvironmental Engineering* 129(11):1054
- Castejón Limas M, Ordieres Meré JB, González Marcos A, de Pisón Ascacibar FJM, Pernía Espinoza AV, Alba Elías F (2010) AMORE: A MORE flexible neural network package. R package version 0.2-12. <http://CRAN.R-project.org/package=AMORE>
- Coppola EA, Rana AJ, Poulton MM, Szidarovszky F, Uhl VW (2005) A neural network model for predicting aquifer water level elevations. *Ground Water* 43(2):231–241
- Cronican A, Gribb M (2004) Literature review: Equations for predicting hydraulic conductivity based on grain-size data. Supplement to technical note entitled: Hydraulic conductivity prediction for sandy soils. *Ground Water* 42(3):459–464
- Darcy H (1856) *Les Fontaines Publiques de la Ville de Dijon*. Dalmont, Paris
- Fletcher L, Katkovnik V, Steffens F, Engelbrecht A (1998) Optimizing the number of hidden nodes of a feedforward artificial neural network. In: *Neural networks proceedings IEEE world congress on computational intelligence, the 1998 IEEE international joint conference*, vol 2, pp 1608–1612
- Gunst FR, Mason LR (1980) *Regression analysis and its applications: A data oriented approach*. Dekker, New York, 402 pp
- Haykin S (1999) *Neural networks—A comprehensive foundation*, 2nd edn. Prentice Hall, New Jersey
- Hazen A (1892) Some physical properties of sands and gravels. *Massachusetts State Board of Health Annual Report* 539–556
- Henseler J (1995) Back propagation. In: Braspenning PJ, Thuijsman F, Weijters AJMM (eds) *Artificial neural networks; an introduction to ANN theory and practice*, vol 10. Springer, Berlin, pp 37–66
- Hill T, Lewicki P (2007) *STATISTICS: Methods and applications*. StatSoft, Tulsa
- Hornik K, Stinchcombe M, White H (1989) Multilayer feedforward networks are universal approximators. *Neural Netw* 2(5):359–366
- Hornik K (1991) Approximation capabilities of multilayer feedforward networks. *Neural Netw* 4(2):251–257
- Jain A, Kumar AM (2007) Hybrid neural network models for hydrologic time series forecasting. *Appl Soft Comput* 7(2):585–592
- Joorabchi A, Zhang H, Blumenstein M (2009) Application of artificial neural networks to groundwater dynamics in coastal aquifers. In: *Proceedings of the 10th international coastal symposium*. *J Coast Res* 56:966–970
- Khalil B, Ouarda TBMJ, St-Hilaire A (2011) Estimation of water quality characteristics at ungauged sites using artificial neural networks and canonical correlation analysis. *J Hydrol* 405(3–4):277–287

- Kişî Ö (2007) Streamflow forecasting using different artificial neural network algorithms. *J Hydrol Eng* (October):532–539
- Kleinbaum DG, Kupper LL, Muller KE (2007) Applied regression analysis and other multivariable methods. Cengage learning, 906 pp
- Klute A (1965) Laboratory measurements of hydraulic conductivity of saturated soil. In: Black CA et al (eds) *Methods of soil analysis. Part 1, Agronomy*, vol 9, pp 210–220
- Kozeny J (1927) Ueber kapillare Leitung des Wassers im Boden. *Sitzungsber Akad Wiss Wien* 136(2a):271
- Linderman M, Liu J, Qi J, An L, Ouyang Z, Yang J, Tan Y (2004) Using artificial neural networks to map the spatial distribution of understorey bamboo from remote sensing data. *Int J Remote Sens* 25(9):1685–1700
- Minasny B, Hopmans JW, Harter T, Eching SO, Tuli A, Denton MA (2004) Neural networks prediction of soil hydraulic functions for alluvial soils using multistep outflow data. *Soil Sci Soc Am J* 68(2):417–429
- Morshed J, Kaluarachchi JJ (1998) Application of artificial neural network and genetic algorithm in flow and transport simulations. *Adv Water Resour* 22(2):145–158
- Nakhaei M (2005) Estimating the saturated hydraulic conductivity of granular material using artificial neural network based on grain size distribution curve. *J Sci Islam Repub Iran* 16(1):55–62
- Pachepsky YA, Rawls WJ, Timlin DJ (1999) The current status of pedotransfer functions: Their accuracy reliability and utility in field- and regional-scale modeling. In: Corwin DL, Loague K, Ellsworth TR (eds) *Assessment of non-point source pollution in the vadose zone: Geophysical monograph*, vol 108. American Geophysical Union, Washington, pp 223–234
- Prechelt L (1998) Automatic early stopping using cross validation: Quantifying the criteria. *Neural Netw* 11(4):761–767
- R Development Core Team (2010) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. <http://www.R-project.org/>
- Sarkar D (1995) Methods to speed up error back-propagation learning algorithm. *ACM Comput Surv* 27(4):519–542
- Sarle WS (ed) (1997) Neural Network FAQ. Periodic posting to the Usenet newsgroup comp.ai.neural-nets. <ftp://ftp.sas.com/pub/neural/FAQ.html>, visited on 11/11/2010
- Schaap M, Leij FJ (1998) Using neural networks to predict soil water retention and soil hydraulic conductivity. *Soil Tillage Res* 47(1–2):37–42
- Schaap MG, Leij FJ, van Genuchten MT (1998) Neural network analysis for hierarchical prediction of soil hydraulic properties. *Soil Sci Soc Am J* 62:847–855
- Schaap M, Leij F, Van Genuchten MTh (2001) Rosetta: A computer program for estimating soil hydraulic parameters with hierarchical pedotransfer functions. *J Hydrol* 251(3–4):163–176
- Scott D (1979) On optimal and data-based histograms. *Biometrika* 66:605–610
- Soetens T (2008) Oriënterend bodemonderzoek in het kader van de aankoop van het terrein voor de oppervlakteberging van Umicore. NIRAS/ONDRAF 245.090-MER000, 2008-1119 herz.1
- Stedinger JR, Vogel RM, Lee SU, Batchelder R (2008) Appraisal of the generalized likelihood uncertainty estimation (GLUE) method. *Water Resour Res* 44:W00B06
- Swingler K (1996) Applying neural networks: A practical guide. Academic Press, London
- Tiwari MK, Chatterjee C (2010) Uncertainty assessment and ensemble flood forecasting using bootstrap based artificial neural networks (BANNs). *J Hydrol* 382(1–4):20–33
- Tollenaere T (1990) SuperSAB fast adaptive back propagation with good scaling properties. *Neural Netw* 3(5):561–573
- Valverde Ramírez MC, Campos Velho HF, de Ferreira NJ (2005) Artificial neural network technique for rainfall forecasting applied to the Sao Paulo region. *J Hydrol* 301:146–162
- Van De Genachte G, Mallants D, Ramos J, Deckers JA, Feyen J (1996) Estimating infiltration parameters from basic soil properties. *Hydrol Process* 10(5):687–701
- Vienken T, Dietrich P (2011) Field evaluation of methods for determining hydraulic conductivity from grain-size data. *J Hydrol* 400(1–2):58–71
- Wang T, Wedin D, Zlotnik VA (2009) Field evidence of a negative correlation between saturated hydraulic conductivity and soil carbon in a sandy soil. *Water Resour Res* 45(7):W07503. doi:10.1029/2008WR006865
- Wemaere I, Marivoet J, Labat S (2008) Hydraulic conductivity variability of the Boom clay in north-east Belgium based on four core drilled boreholes. *Phys Chem Earth* 33(S1):24–36

- Wemaere I, Marivoet J, Labat S, Beaufays R, Maes T (2002) Mol-1 borehole (April–May 1997): Core manipulations and determination of hydraulic conductivities in the laboratory. SCK•CEN Report R-3590, 56 pp
- Zou R, Lung WS, Wu J (2007) An adaptive neural network embedded genetic algorithm approach for inverse water quality modeling. *Water Resour Res* 43:W08427. doi:[10.1029/2006WR005158](https://doi.org/10.1029/2006WR005158)